

Programa de I+D+I en Ingeniería de Explotación de Información

Informe de Avances Periodo 2011-2016 y Previsión Periodo 2017-2019

Ramón García-Martínez, Hernán Merlino, Darío Rodríguez, Sebastian Martins, Ezequiel Baldizzoni, Santiago Bianco

Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información
Grupo de Ingeniería de Explotación de Información y Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 5533 5600 Ext. 5194
rgm1960@yahoo.com

Resumen

El Programa de I+D+I en Ingeniería de Explotación de Información tiene como objetivo general construir conocimiento aplicable a problemáticas del área sobre los procesos, las metodologías y las técnicas utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en grandes masas de información. El programa evoluciona en la línea de tiempo a partir de los problemas abiertos que se identifican en los proyectos y de los requerimientos que plantea el sistema productivo regional.

Palabras clave: Minería de datos, Explotación de Información, Ingeniería de Procesos, Ingeniería de Explotación de Información.

Contexto

El Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús, realizó un ordenamiento de las actividades de I+D+I, estructurándolas en Líneas de Investigación [DDPyT, 2016]. Cada Línea de Investigación se estructura en Programas de I+D+I. En particular, la Línea de Investigación en el Campo de Productividad en Informática se estructura en los siguientes Programas de I+D+I: Ingeniería de Sistemas de Información, Ingeniería de Sistemas Inteligentes, Ingeniería de Explotación de Información, Ingeniería de Espacios Virtuales de Trabajo, e Ingeniería de Desarrollo de Sistemas Especiales.

Introducción: Resultados Periodo 2011-2016

En el Periodo 2011-2016 se acreditaron y desarrollaron los siguientes Proyectos de I+D+I:

- Ingeniería de Proyectos de Explotación de Información para PyMEs (Código: 33A105, Periodo:2011-2012)
- Aseguramiento de la Calidad para Proyectos de Explotación de Información (Código: 33B102, Periodo 2012-2013)
- Instrumentos para la Gestión de Proyectos de Explotación de Información (Código: 33A167, Periodo:2013-2014)
- Procesos de Explotación de Información Georeferenciada (Código: 33B133, Periodo 2015-2016)
- Modelos de Proceso para Ingeniería de Explotación de Información para PYMES: Abordaje Ágil y Abordaje Robusto (Código: 33A205, Periodo 2015-2016)

En el Proyecto UNLa 33A105 *Ingeniería de Proyectos de Explotación de Información para PyMEs* [GISI-UNLa, 2011] se desarrollaron los siguientes artefactos de Ingeniería de Proyectos de Explotación de Información: [a] una batería de técnicas de educación y formalismos de documentación de requerimientos; [b] la versión preliminar de un modelo de procesos y se comenzó a explorar las métricas asociadas; [c] un modelo de ciclo de vida; y [d] un mapa de actividades. En el Proyecto UNLa 33B102 *Aseguramiento de la Calidad para Proyectos de Explotación de Información* [GISI-UNLa, 2012] se revisaron los paradigmas existentes sobre calidad en proyectos de software aplicables a proyectos de explotación de información, y se planteó su adaptación para lograr el Aseguramiento de la Calidad en este tipo de Proyectos.

En el Proyecto UNLa 33A167 *Instrumentos para la Gestión de Proyectos de Explotación de Información* [GISI-UNLa, 2013] se desarrollaron mediante la metodología de prototipado evolutivo, los siguientes instrumentos para el control y la gestión de Proyectos

de Explotación de Información para PyMEs: [a] test de viabilidad, [b] procedimiento de derivación del problema de negocio en problema de explotación de información, [c] formalismos para documentación de requisitos, [d] técnicas y métodos de aseguramiento de la calidad del proceso y del producto, y [e] método de estimación de recursos.

En el Proyecto UNLa 33B133 *Procesos de Explotación de Información Georeferenciada* [GISI- UNLa, 2014] se sentaron las bases para caracterizar los procesos de explotación de información georeferenciada asociados a detección de patrones y regularidades importantes en artefactos de información geográficos, y se identificaron las tecnologías de Sistemas Inteligentes que se pueden utilizar para los procesos caracterizados. En el Proyecto UNLa 33A205 *Modelos de Proceso para Ingeniería de Explotación de Información para PYMES: Abordaje Ágil y Abordaje Robusto* [GISI-UNLa, 2015] se desarrolló un modelo de proceso que propone fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; y que soporta de forma completa e integrada las actividades de administración y desarrollo de proyectos en dos versiones, una orientada a microproyectos y otra orientada a proyectos regulares.

En el Periodo 2011-2016 en el marco de los Proyectos de I+D+I que integraban el Programa se desarrollaron y defendieron las siguientes tesis doctorales y de maestría:

- Procedimientos de Explotación de Información en Identificación de Datos Faltantes, con Ruido o Inconsistentes (Tesis Doctoral Universidad de Malaga, 2014)
- Viabilidad y Estimación de Proyectos de Explotación de Información (Tesis Doctoral Universidad Nacional de La Plata, 2014)
- Gestión de la Educación Superior en Contextos de Masividad Basada en Tecnologías Inteligentes de Transformación de Información (Tesis Doctoral Universidad Nacional de Córdoba, Defensa pendiente)
- Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramientas de Validación (Tesis de Magister Universidad Politécnica de Madrid/ ITBA, 2011)
- Un Modelo de Procesos para Proyectos de Explotación de Información (Tesis de Magister UTN-FRBA, 2012)

- Descubrimiento de Conocimiento Mediante la Integración de Algoritmos de Explotación de la Información (Tesis de Magister UTN-FRBA, 2012)
- Propuesta de Métricas para Proyectos de Explotación de Información (Tesis de Magister UTN-FRBA, 2014)
- Proceso de Conceptualización del Entendimiento del Negocio para Proyectos de Explotación de Información (Tesis de Magister UTN-FRBA, 2014)
- Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos (Tesis de Magister Universidad Nacional de La Plata, 2016)

En la Tesis *Procedimientos de Explotación de Información en Identificación de Datos Faltantes, con Ruido o Inconsistentes* [Kuna, 2014] se establece una taxonomía relacionada con los métodos, técnicas y algoritmos de detección de valores anómalos en bases de datos, analizando las ventajas y desventajas de cada una de ellos, y se proponen procedimientos de explotación de información que combinados entre sí permitan detectar los campos que tienen valores atípicos en bases de datos, combinando distintas técnicas de minería de datos, entre ellas específicamente las de clustering, algoritmos de inducción, redes bayesianas, principios de la teoría de la Información, entre otras, para lograr mejorar de esa manera la calidad de los datos.

En la Tesis *Viabilidad y Estimación de Proyectos de Explotación de Información* [Pytel, 2014] se proponen dos procesos ad-hoc que permitan en el marco de los Proyectos de Explotación de Información: [a] evaluar la viabilidad de a partir de la información disponible al comienzo del proyecto, y [b] estimar los recursos necesarios con particular énfasis en su utilización en PyMEs.

En la Tesis *Gestión de la Educación Superior en Contextos de Masividad Basada en Tecnologías Inteligentes de Transformación de Información* [Díaz, 2017] propone el uso de Tecnologías Inteligentes de Explotación de la Información y otras herramientas del campo de la Inteligencia Computacional para construir categorías y enriquecer teorías a la luz del Paradigma Interpretativo de investigación en las Ciencias Sociales. En particular, a partir de la experiencia piloto, se proporcionará un mejor conocimiento del estudiante de cursos masivos de esta Universidad en sus características socio- económicas y de rendimiento académico.

En la Tesis *Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramientas de Validación* [Pytel, 2011] se propone una herramienta de software que ayude a validar el método de estimación de esfuerzo para proyectos de Explotación de Información denominado Método Matemático Paramétrico de Estimación para Proyectos de Data Mining (DMCoMo). La herramienta permite generar un banco de pruebas con las características de proyectos de Explotación de Información relevantes para el método de estimación, y así luego analizar el comportamiento del método de estimación seleccionado.

En la Tesis *Un Modelo de Procesos para Proyectos de Explotación de Información* (Tesis de Magister UTN-FRBA, 2012) Se busca desarrollar un modelo de procesos para proyectos de explotación de información, atendiendo los siguientes objetivos específicos: [a] Establecer diferencias y relaciones entre los proyectos de explotación de información y los proyectos clásicos, [b] Determinar límites, alcance y componentes del modelo MoProSoft aplicables a Proyectos de Explotación de Información, [c] Identificar los procesos específicos de los proyectos de explotación de información a través de las distintas metodologías existentes que se aplican a dichos proyectos y que no estén cubiertos por los modelos de procesos de software considerados, [d] Seleccionar los procesos a incluir en el nuevo modelo, y [e] Desarrollar el nuevo modelo basándose en MoProSoft a partir de la documentación de los procesos identificados y utilizando el patrón de procesos especificado por MoProSoft.

En la Tesis *Descubrimiento de Conocimiento Mediante la Integración de Algoritmos de Explotación de la Información* [Lopez Nocera, 2012], sobre la premisa que en Minería de Datos se parte del supuesto que los dominios son equivalentes o directamente indistinguibles entre sí a los efectos experimentales; se postula que esto no es así, y que la caracterización de los dominios en uso durante la fase experimental al estudiar algoritmos de minería de datos tiene amplia injerencia en los resultados observados y finalmente obtenidos. La postulación es corroborada con resultados experimentales.

En la Tesis *Propuesta de Métricas para Proyectos de Explotación de Información* [Basso, 2014] se parte de la premisa que las métricas usuales de la Ingeniería de Software y la Ingeniería del Conocimiento no son adecuados para los proyectos de explotación de información, ya que las etapas de desarrollo y los

parámetros utilizados son de naturaleza y características diferentes. En ese contexto, se propone un conjunto de métricas aplicables al desarrollo de un proyecto de Explotación de Información para PyMEs, centrado en el Modelo MoProPEI.

En la Tesis *Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos* [Cigliuti, 2016] se propone el uso de procesos de explotación de información para identificar el comportamiento de comunidades educativas basado en resultados académicos.

Objetivos: Líneas de Trabajo, Desarrollo e Innovación Periodo 2017-2019

Para el periodo 2017-2019 las Líneas de Trabajo, Desarrollo e Innovación se encuentran articuladas a través de los siguientes Proyectos de I+D+I que, ha la fecha de esta comunicación, cuentan con acreditación institucional y han comenzado su desarrollo:

- Integración de Herramientas y Procesos de Ingeniería de Explotación de Información (Código: AH2016 33A 80020160400001LA, Periodo: 2017- 2018)
- Identificación de Patrones de Adicción a Sustancias Psicoactivas Basada en Ingeniería de Explotación de Información (Código: AH2016 33A 800201605000 02LA, Periodo: 2017-2018)
- Ingeniería de Proyectos de Explotación de Grandes Volúmenes de Datos (Código PIO UNLa-CONICET 22420160100032CO, Periodo: 2017-2018)

En el Proyecto *Integración de Herramientas y Procesos de Ingeniería de Explotación de Información*, se busca desarrollar un marco integrado de herramientas conceptuales y procesos de utilizables en producción para llevar adelante las actividades de Ingeniería de un Proyecto de Explotación de Información. El marco propuesto se orienta a contribuir a que los profesionales informáticos dispongan de guías que ordenen el trabajo y permitan establecer y ofrecer hitos para los entregables de proyectos en el área de ingeniería de explotación de información orientadas a inteligencia de negocio.

En el Proyecto *Identificación de Patrones de Adicción a Sustancias Psicoactivas Basada en Ingeniería de Explotación de Información*, se busca explorar la utilización de Ingeniería de Explotación de Información como abordaje alternativo a la identificación de patrones de consumo de sustancias

psicoactivas tomando como insumo la información disponible en encuestas del INDEC con énfasis en la identificación de [a] variables descriptoras concurrentes con el consumo de sustancias psicoactivas, y [b] patrones de consumo y de consumidores de sustancias psicoactivas describibles en términos de instancias de las variables concurrentes con el consumo de sustancias psicoactivas.

En el Proyecto *Ingeniería de Proyectos de Explotación de Grandes Volúmenes de Datos* se busca desarrollar el cuerpo de conocimiento necesario para las actividades tempranas de Ingeniería de Proyectos de Explotación en Grandes Volúmenes de Datos (PEGVD) con focalización en su transferencia al sector PyMEs de Informática. El proyecto busca proveer un marco integrado de trabajo para PEGVD que contempla generar la versión de producción de las siguientes herramientas: [a] un modelo de procesos de decisión estratégica, control y gestión y desarrollo para PEGVD, [b] una batería de técnicas de educación y formalismos de documentación de requerimientos para PEGVD, [c] un test de viabilidad para PEGVD, [d] un modelo de ciclo de vida para PEGVD, [e] un mapa de actividades para PEGVD, [f] un método de estimación de recursos para PEGVD, [g] un procedimiento de derivación del problema de negocio en problema de exploración en grandes volúmenes de datos, y [h] un conjunto de técnicas y métodos de aseguramiento de la calidad del proceso y del producto para PEGVD.

En el Periodo 2017-2019 en el marco de los Proyectos de I+D+I que integran el Programa se prevé desarrollar y finalizar las siguientes tesis doctorales y de maestría:

- Integración de Métodos en Ingeniería de Explotación de Información (Tesis Doctoral Universidad Nacional de La Plata, Finaliza 2017)
- Metodología para la Especificación de Requisitos en Proyectos de Explotación de Información (Tesis Doctoral Universidad Nacional de La Plata, Finaliza 2018)
- Integración de Métodos de Descubrimiento de Conocimiento Embebido en Fuentes de Información Desestructuradas (Tesis Doctoral Universidad Nacional de La Plata, Finaliza 2019)
- Formulación de Procesos para una Ingeniería de Explotación de Información Espacial (Tesis Doctoral Universidad Nacional de La Plata, Finaliza 2020)

- Propuesta de Proceso para Ingeniería de Explotación de Información Centrado en Métodos y Técnicas - Núcleo IMMAT (Tesis de Magister UTN-FRBA, Finaliza 2018)

En la Tesis *Integración de Métodos en Ingeniería de Explotación de Información* [Martins, 2014] se busca sistematizar el conocimiento existente sobre ingeniería de explotación de información y formular una propuesta unificada de modelo de proceso que integre todos los métodos, técnicas y herramientas de administración y desarrollo de proyectos en esta sub-disciplina informática.

En la Tesis *Metodología para la Especificación de Requisitos en Proyectos de Explotación de Información* [Pollo-Cattaneo, 2010] se busca definir una metodología de educación de requisitos para proyectos de explotación de información en el marco del ciclo de vida utilizado en este tipo de proyectos. La tesis se enfocará a construir una metodología que permita la educación de requisitos para proyectos de explotación de información teniendo presente los distintos procesos que involucra el ciclo de vida usado en este tipo de proyectos. Se considera que la tesis implica realizar investigación básica en la definición del proceso de educación de requisitos en proyectos de explotación de información.

En la Tesis *Integración de Métodos de Descubrimiento de Conocimiento Embebido en Fuentes de Información Desestructuradas* se busca construir una familia de métodos de extracción de conocimiento tal que dado una estructura de información inicial como entrada, que contienen conocimiento embebido, genere un conjunto de piezas de conocimiento (reglas de producción, subgrafos de una red semántica, entre otros). Tiene los siguientes Objetivos Específicos: (i) Realizar una comparación entre los distintos métodos de extracción de conocimiento relevados en la literatura, indicando, para diversas estructuras de información, la calidad y los tiempos de ejecución asociados a cada uno; (ii) Identificar fortalezas y debilidades en los diversos métodos evaluados en OE1 con el fin de detectar las condiciones particulares bajo las cuales es conveniente utilizar un método por sobre otro y (iii) Desarrollar una familia de métodos integrados de extracción de conocimiento que exhiban un mejor comportamiento que los métodos individuales integrados.

En la Tesis *Formulación de Procesos para una Ingeniería de Explotación de Información Espacial*, dadas las particularidades de la Explotación de

Información Espacial en contraste con la Explotación de Información convencional, resulta de interés proponer un conjunto de Procesos de Explotación de Información Espacial relevantes para la inteligencia de Negocio. Estos Procesos de Explotación de Información Espacial deberán incluir las fuentes de Información necesarias, tratamiento de los datos y los algoritmos o familias de algoritmos de Explotación de Información a ser utilizados para el descubrimiento de Patrones en la Información Espacial.

En la Tesis *Propuesta de Proceso para Ingeniería de Explotación de Información Centrado en Métodos y Técnicas - Núcleo IMMAT*, se busca desarrollar la adaptación del núcleo SEMAT, que fue pensado para proyectos de Ingeniería de Software, a proyectos de Ingeniería de Explotación de Información con el fin de aportar a esta última, buenas practicas que colaboren en obtener los mejores resultados, de mejor calidad y optimizando los recursos.

Formación de Recursos Humanos

En este Programa participan 3 investigadores formados, y un investigador formado adscripto. A la fecha de esta comunicación se han radicado en el Programa: 7 tesis doctorales (4 defendidas), 7 tesis de magíster (6 defendidas), y 8 Trabajos Finales de Licenciatura (5 defendidos).

Referencias

- Baldizzoni, E. 2016. *Investigación en Progreso: Propuesta de Proceso para Ingeniería de Explotación de Información Centrado en Métodos y Técnicas - Núcleo IMMAT*. Revista Latinoamericana de Ingeniería de Software 4(1): 55-58. ISSN 2314-2642.
- Basso, D. 2014. *Propuesta de Métricas para Proyectos de Explotación de Información*. Tesis de Magister en Ingeniería de Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
- Cigliuti, P. 2016. *Proceso de Identificación de Comportamiento de Comunidades Educativas basado en Resultados Académicos*. Tesis de Magister en Tecnología Informática Aplicada en Educación. Facultad de Informática. UNLP.
- DDPyT, 2016. *Líneas de Investigación - Programas I+D+I - Grupos de I&D y Producción Científica*. <http://ddpyt.unla.edu.ar/documentos/DDPyT-Lineas-IyD.pdf>
- Díaz, L. 2014. *Gestión de la Educación Superior en Contextos de Masividad Basada en Tecnologías Inteligentes de Transformación de Información*. Tesis Doctoral en Administración Pública en Desarrollo. Universidad Nacional de Córdoba. <http://sistemas.unla.edu.ar/sistemas/gisi/GISI-TD-2014-02.htm>
- GISI-UNLa, 2009. *Proyecto UNLa 33A081 Sistemas de Información para Inteligencia de Negocio*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33A081.htm>
- GISI-UNLa, 2011. *Proyecto UNLa 33A105 Ingeniería de Proyectos de Explotación de Información para PyMEs*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33A105.htm>
- GISI-UNLa, 2012. *Proyecto UNLa 33B102: Aseguramiento de la Calidad para Proyectos de Explotación de Información*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33B102.htm>
- GISI-UNLa, 2013. *Proyecto UNLa 33A167 Instrumentos para la Gestión de Proyectos de Explotación de Información*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33A167.htm>
- GISI-UNLa, 2014. *Proyecto UNLa 33B133 Procesos de Explotación de Información Georeferenciada*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33B133.htm>
- GISI-UNLa, 2015. *Proyecto UNLa 33A205 Modelos de Proceso para Ingeniería de Explotación de Información para PYMES: Abordaje Ágil y Abordaje Robusto*. <http://sistemas.unla.edu.ar/sistemas/gisi/UNLa-33A205.htm>
- Kuna, H. 2014. *Procedimientos de Explotación de Información para la Identificación de Datos Faltantes con Ruido e Inconsistentes*. Tesis Doctoral en Ingeniería de Sistemas y Computación. Universidad de Málaga.
- Lopez Nocera, M. 2012. *Descubrimiento de Conocimiento Mediante la Integración de Algoritmos de Explotación de la Información*. Tesis de Magister en Ingeniería de Sistemas de Información. UTN-FRBA.
- Martins, S. 2014. *Integración de Métodos en Ingeniería de Explotación de Información*. Tesis Doctoral en Ciencias informáticas en Desarrollo. Universidad Nacional de La Plata. <http://sistemas.unla.edu.ar/sistemas/gisi/GISI-TD-2014-01.htm>
- Peralta, F. 2014. *Proceso de Conceptualización del Entendimiento del Negocio para Proyectos de Explotación de Información*. Tesis de Magister en Ingeniería de Sistemas de Información. UTN-FRBA.
- Pollo-Cattaneo, F. 2010. *Metodología para la Especificación de Requisitos en Proyectos de Explotación de Información*. Tesis Doctoral en Ciencias informáticas en Desarrollo. Universidad Nacional de La Plata. <http://sistemas.unla.edu.ar/sistemas/gisi/GISI-TD-2010-02.htm>
- Pytel, P. 2011. *Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramienta Para Su Validación*. Tesis de Magister en Ingeniería del Software. Convenio Universidad Politécnica de Madrid e Instituto Tecnológico Buenos Aires.
- Pytel, P. 2014. *Viabilidad y Estimación de Proyectos de Explotación de Información*. Tesis Doctoral en Ciencias informáticas. Universidad Nacional de La Plata.
- Vanrell, J. 2012. *Un Modelo de Procesos para Proyectos de Explotación de Información*. Tesis de Magister en Ingeniería de Sistemas de Información. UTN-FRBA.

Nota: Se puede acceder a la Producción Científica del Programa a través de la página web del Grupo de Ingeniería en Explotación de Información:

<http://sistemas.unla.edu.ar/sistemas/gisi/GIEL.html>